DOCUMENT RESUME

ED 252 459 SO 016 111

AUTHOR Johnson, Grant
TITLE When Law-Related Education is a Deterrent to Delinquency: Evaluation Methods and Findings and Addendum: A Preliminary Response to James P. Shaver's Paper, "The Law-Related Education Evaluation Project: A Methodological Critique of the 'Impacts on Students' Findings."
INSTITUTION Center for Action Research, Inc., Boulder, Colo.
PUB DATE Apr 84
NOTE 20p.; Prepared as part of the "LRE Evaluation Project: A Symposium on Methods, Findings, Conclusions," at the Rocky Mountain Regional Conference of the National Council for the Social Studies (Phoenix, AZ, April 11-14, 1984). For related documents, see ED 247 164 and SO 016 110 and SO 016 112.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Delinquency; *Educational Assessment; Educational Research; Followup Studies; Legal Education; *Outcomes of Education; Program Effectiveness; *Reliability; Research Design; Research Methodology; Research Problems; Secondary Education; Student Behavior; *Validity
IDENTIFIERS *Law Related Education; Shaver (James P)

ABSTRACT

The methodology and findings of the Law-Related Education Project's research reports which evaluated the effects on secondary students of law-related education (LRE), are discussed. A response to James P. Shaver's negative critique of the Law-Related Education Project's research studies also is provided. A detailed account of the research results obtained from one site--Bill Reed Junior High School (Loveland, Colorado)--that demonstrated significant favorable LRE impact is included. Through student questionnaires, the original study compared the delinquency behavior of ninth-graders enrolled in LRE classes with the behavior of students enrolled in civics classes. To determine long-term effects of LRE, a follow-up study was also conducted. The general summation of findings was that when properly implemented, LRE can serve as a deterrent to delinquent behavior. Effective LRE instruction must consist of certain features and teaching methods; not all LRE classes deter delinquent behavior. Shaver's critique does not give cause to revise any conclusions regarding the research. A defensible critique must present accurately the research that it examines. Shaver's did not; the conclusions in his review rely on very selective choices of data. With respect to the appropriateness of relying on student self-reports in measuring delinquency, many studies have demonstrated that self-reports are valid and reliable. Completely ignored by the critique was the fact that the test of the theory of delinquency causation yielded overwhelmingly supportive findings. (RM)

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Grant Johnson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED252459

WHEN LAW-RELATED EDUCATION IS A DETERRENT TO DELINQUENCY: EVALUATION METHODS AND FINDINGS

and

ADDENDUM: A PRELIMINARY RESPONSE TO JAMES P. SHAVER'S PAPER

Grant Johnson
Center for Action Research, Inc.

April 1984

SP 016 111

WHEN LAW-RELATED EDUCATION IS A DETERRENT TO DELINQUENCY: EVALUATION METHODS AND FINDINGS

by

Grant Johnson
Center for Action Research, Inc.
1125 Spruce Street
Boulder, CO 80302

Prepared as part of

THE LRE EVALUATION PROJECT: A SYMPOSIUM ON METHODS, FINDINGS, CONCLUSIONS

Rocky Mountain Regional Conference
of the
National Council for the Social Studies

Phoenix, Arizona
April 11-14, 1984

OVERVIEW

Evaluation of the effects on students of law-related education (LRE) occurred in spring of 1981, fall of 1982, and fall of 1983. The research included questionnaires administered to students at the start and end of semesters, use of comparison classes, and classroom observations. Sixty-nine classes in thirty-six schools were in the study.

In successive years, both the quality of the courses evaluated and the quality of the research improved. The general summation of findings remained the same each year: *When properly implemented, LRE can serve as a deterrent to delinquent behavior.* Better courses and better research make this perennial conclusion more defensible today than it was two years ago, and the features that constitute "proper implementation" now can be specified more precisely.*

To say that "LRE reduces delinquency" seriously misstates the conclusion of the study. Today as in 1981, very diverse courses and events bear the name "LRE." Even classes having identical course descriptions may differ drastically in practice. A belief that everything called LRE will have a uniform effect on behavior (or anything else) is unreasonable. From the standpoint of delinquency prevention, some LRE classes in the study clearly demonstrated significant favorable impact. Less powerfully (due to less stringent design), the research also identified both ineffective LRE classes and harmful ones.

What follows is a brief review of methods and findings from the entire study, then a more detailed account of results obtained from nine classes at the site having the strongest research design and in several respects the most ambitious program in the study.

METHODS

With the aim of adding to the prospects for giving credible and useful advice (formative evaluation) and as a means of addressing threats to the validity of the findings, evaluators combined student impact

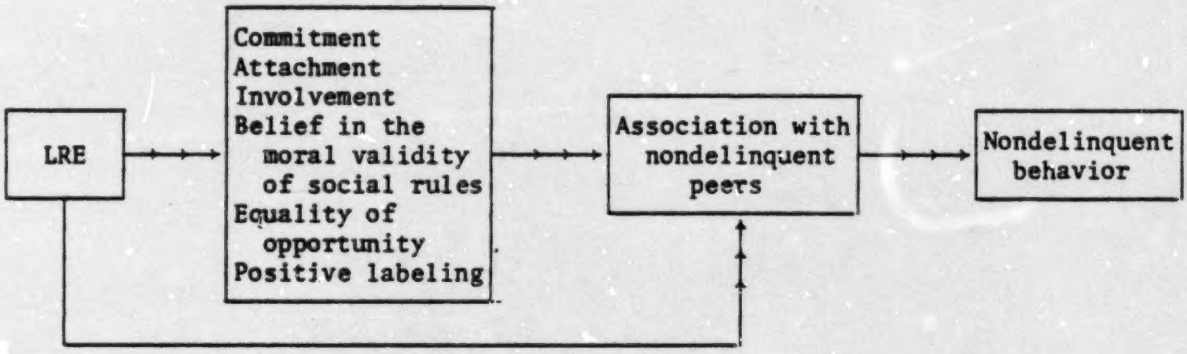
---

*Mary Jane Turner's paper, "An Evaluation of Law-Related Education: Implications for Teaching," describes those features.

testing with interviews, questionnaires, and direct observations involving a broad range of participants. Members of the evaluation team have interviewed teachers, school administrators, community resource people, and others who have participated in the classroom. They have observed in experimental and comparison classrooms, in training sessions, and in district seminars. Measures have included high inference scales and low inference observation formats.

Grounds for expecting LRE to reduce delinquent behavior come from social control, labeling, and strain theories. These theories contain a number of elements established by previous research as causally related to law-abiding behavior; the elements appear in the simplified diagram below.

Figure 1: Potential Links Between LRE and Behavior



Most of the outcome measures used in the student impact tests are operationalizations of the elements in the diagram. Very briefly, *commitment* refers to a valued stake in conventional activities that could be lost as a result of rule-breaking. A second control process is *attachment* to other people. To violate a norm is to violate the wishes and expectations of others; a low level of attachment makes violations more likely. A third process is *involvement*, a person's present investment of time and energy in certain conventional activities (excluding sports and passive entertainment). A fourth control process is *belief* that rules governing behavior are fair enough to be worth obeying consistently and that police activities and judicial mechanisms used to enforce the rules are fair as well. *Equality of opportunity* refers to a perception that legitimate avenues for achieving goals are about as open to you as to

others. *Positive labeling* refers to an assessment that other people who are important in your life regard you as mostly law-abiding and productive. The greater the degree to which these elements are present, the less a young person's need to obtain personal rewards from delinquent peers and the less the likelihood that delinquent behavior will ensue.

In addition to scales derived from the theoretical model, the student impact questionnaires included:

A test of knowledge about the law and principles underlying the legal system, varied to suit the particular curriculum in use at each school.

A series of twenty-one items to assess the frequency with which a subject had committed each of eleven types of offense during the preceding semester.

Items allowing students to rate their course.

At one site in 1982 and 1983 (for a total of twelve LRE classes), the evaluators obtained random assignment of students between LRE and control classes. In the closest approximation to a truly experimental design that the evaluators were able to negotiate elsewhere, the project called for the selection of experimental and comparison classes of comparable age, sex, ethnic, and academic characteristics in each site. The students in each class at these sites were, however, assigned by the usual student placement process in their respective schools, and were neither randomly assigned nor matched. Problems associated with imperfect match led the evaluators (in 1981) to perform separate analyses for male and female students and (in 1982 and 1983) to pretest more classes than they intended to study, deleting at the beginning of the semester the pairs of classes in which experimental and control subjects were least comparable.

In analyzing longitudinal data from the questionnaires in 1981, residual gain analysis was used to control for the effect of time-1 (pretest) scale scores and behavior frequencies on changes during the semesters studied. To assess differences between an LRE and a control class, mean residual gains for all students in the respective classes were compared; t-tests were applied to determine which differences of means were statistically significant. In 1982 and 1983, multiple regression analysis was used to control for the effects of both time-1 scores and age. In the analysis, the time-2 score or frequency was treated as the dependent variable, with time-1 score

and age specified as the first two independent variables to enter the analysis. Only then did the LRE/nonLRE variable enter the equation. As a consequence, the analysis showed how much additional effect LRE had, over and above the effects of pretest score and age. For measures administered only at time-2 (e.g., rating of the course just completed), t-tests were used to determine whether the experimental and comparison group means differed significantly.

At the site with a true experimental design, the evaluators conducted a four-month followup study of those in the 1982 study. Experimental and control subjects completed a third questionnaire at the end of the semester following the one in which LRE was offered. Multiple regression analysis was used to determine the degree to which the effects of LRE shown at time-2 still were in evidence four months later.

FINDINGS

The limited classroom study in 1981 produced the first outline of features likely to distinguish successful LRE classes from unsuccessful ones. Seven LRE classes were subjects of both testing and observations. Analysis of each kind of data proceeded independently: until the tallies were complete, neither the person computing quantitative outcomes nor the person assembling information collected by observers was privy to the other's data. Before sharing any information, both staff members ranked the seven classes from most to least effective. One ranking was based on numerical questionnaire results only, the other on observers' predictions of probable success only. In terms of the match between experimental and comparison subjects, number of classes, amount of time spent observing, and the method for handling longitudinal data (residual gain), the 1981 research was the least stringent part of the three year study. Nevertheless, the two independent rankings corresponded for six of the seven classes. The more that observed classroom practices led to a prediction of favorable impact on student attitudes and behavior, the more that such impact was in evidence in the quantitative analysis. Thus, the evaluators were able to suggest certain training and classroom practices likely to be associated with future success of an LRE course.

Three additional LRE classes were subjects of testing only that year, making a total of ten. Despite relatively uniform knowledge gains across

classes, overall results with respect to student attitudes and behavior were a mix of good and bad news. The apparent net effect on these measures was favorable in four classes, unfavorable in four, and undiscernible in the remaining two. The principal value of the findings was formative. Indications of what made some classes successful were the basis for advice on how to make improvements. The shock of learning that LRE did not always perform as hoped gave cause for listening to that advice.

In 1982, the evaluators analyzed quantitative and observational data from thirty LRE classes. In three of these classes (at Bill Reed Junior High School in Colorado), experimental and control subjects had been randomly assigned--creating the first opportunity for a truly experimental design. For students in these classes, there were fourteen statistically significant effects (.05, one-tailed test) out of thirty-two possible. All fourteen effects were in a favorable direction; i.e., reduction of delinquency or improvement in antecedents of nondelinquent behavior among LRE students. In addition, there were nine trends which did not reach statistical significance; eight of the nine also were in a favorable direction. At time-3 followup--four months after completing the course--LRE students maintained or increased their advantage over control subjects with respect to twelve of the fourteen significant effects and six of the eight favorable trends (as well as maintaining the one unfavorable trend).

At the remaining twenty-seven sites, both the significant differences between LRE and comparison classes and nonsignificant trends suggest that these classes, too, were effective in more ways that not; lack of strong equivalence at some sites, however, leaves these latter findings open to question. In the multiple regression procedure used, effects that could be due either to LRE or to time-1 differences between experimental and comparison groups are attributed to the latter. Where time-1 differences between the groups are substantial on a given measure, the prospects that LRE will demonstrate an effect (either favorable or unfavorable) are small.

The 1982 findings permitted further progress in the formative evaluation. Features associated with effectiveness became more sharply defined and were used in training of teachers in the 1983 study. At Bill Reed Junior High School, that training included police officers who would participate regularly in classroom instruction. The following section describes the 1983 evaluation at that school.

EVALUATION OF LAW-RELATED EDUCATION AT BILL REED JUNIOR HIGH SCHOOL, FALL 1983

Bill Reed Junior High School (Loveland, Colorado) began offering law-related education (LRE) on an experimental basis in 1982. Strong evidence that year of favorable impact on students led to an expansion of the program; in the fall of 1983, four-fifths of the ninth-graders received this course of instruction. The course again had substantial favorable effects, demonstrating the power of properly implemented LRE as a school-based delinquency prevention program.

Evaluation Design. Extraordinary cooperation received from school administrators and staff permitted the evaluators to use a true experimental design to assess impact. Before the start of the semester, an evaluation staff person applied a table of random numbers to the roster of ninth-graders to determine the assignment of each student to one of the three courses: LRE taught by Instructor A (five sections), LRE taught by Instructor B (four sections), or traditional civics without LRE (two sections). Students assigned to traditional civics were the control group.

Random assignment between experimental (LRE) and control classes avoided a research weakness encountered in other schools in the national study, where either steering by counselors of certain students into LRE or (in the case of elective courses) self-selection could result in something other than a representative cross-section of students enrolling in LRE. Random assignment between the two experimental groups made it possible to assess the effects of any unplanned differences between the two LRE teachers' courses.

The evaluators obtained matched pre- and posttest data from 229 LRE students and 43 control subjects. Besides assessing impact, the evaluators monitored the conduct of the classes. Previous research had identified several features of LRE associated with favorable impact on students. The evaluators obtained information from two sources to show the degree to which these features were present and the form they took in the classes at Bill Reed Junior High School. First, evaluation staff observed fifty-six class periods during the semester. Second, the LRE teachers submitted reports on what occurred in their classes each day.

Implementation Design. For several years, Bill Reed Junior High School has trained its teachers in innovative strategies and encouraged their use in the classroom. The LRE and control teachers in this study all were skilled in techniques to promote high student involvement and interaction in the learning process, mastery learning, and student team learning. With general quality of instruction held constant across classes, differences in outcomes between experimental and control students would represent the unique impact of LRE as a subject over and above the impact of superior instructional techniques.

The principal distinguishing characteristics of LRE were the content covered, activities associated with that content, and use of outside resource persons (police officers) in the classroom. The activities included mock trials, police ridealongs, home security audits, and students' taking the roles of police and other professionals in the justice system.

Through negotiations with the Loveland chief of police, the school principal obtained commitment of eight officers to the LRE program for the entire semester. This made it possible for at least one officer to participate three days a week in each LRE class. Teachers and police officers received three days of joint training prior to the start of the semester. Subsequently, the officers were able to assist in instruction by developing lesson plans; presenting topics such as search and seizure, DUI, preventing sexual assault, and being a good witness; interacting with students and teachers around controversial issues; and providing first-hand knowledge related to textbook content.

Data Analysis. Although random assignment arguably made it less critical, multiple regression analysis (as described earlier) was used to analyze the longitudinal data. Because of the control for pretest score and age, the effect of LRE (B-weight) shown for some measures is more conservative than would be obtained by simply subtracting the control group mean from the experimental mean at time-2. For the ten measures administered only at time-2 (those in italics in the table on page 9), a t-test was used to determine whether the experimental and control group means differed significantly.

Impact of LRE on Students. The table on the following page displays the outcome obtained for each measure at Bill Reed Junior High School. Of the forty-one possible effects of LRE that were assessed, there were twenty-four statistically significant favorable effects, another eight nonsignificant favorable trends, and only one unfavorable trend (also nonsignificant). For the remaining eight measures ("zero effect"), differences between LRE students and control subjects were slight enough to have occurred by chance more than 20 times in 100. These findings demonstrate that the LRE program at this school had dramatic favorable impact on the students who participated. Besides imparting significant practical knowledge of legal processes and leaving students with a sense of having completed a superior course, the program reduced their delinquent behavior, improved their attitudes toward school and police, and affected favorably many other factors associated with law-abiding behavior. The impact on students surpasses that obtained at other sites evaluated and even that obtained at this junior high school in the preceding year.

The outcomes shown are for both experimental groups combined. Separate analysis of each LRE teacher's set of classes by itself yields parallel findings for thirty-seven of the forty-one measures. The effects for one experimental group generally mirror the effects for the other. The two groups differed notably, however, on four measures. *Importance of doing well in this class* (measure #6) and *negative labeling by teachers* (#24) showed favorable trends for students taught by Instructor A, but no trends for those taught by Instructor B. *Willingness of students in this class to help one another* (#28) and *frequency of vandalism* (#35) showed favorable effects for students taught by Instructor B, but no effects for those taught by Instructor A. The most striking difference is with respect to vandalism. Instructor B had included two lessons on this topic. In one, students discussed the use of Halloween as a rationalization for damaging property; in the other, students came to recognize vandalism as an indication of disloyalty to their school.

With respect to comparison of 1982 and 1983 findings at this school, one point is of special interest. Although the LRE students again displayed significant gains relative to those in control classes in *favorable attitudes*

BEST COPY AVAILABLE

EFFECTS OF LAW-RELATED EDUCATION AT BILL REED JUNIOR HIGH SCHOOL, FALL 1983

(N= 229 LRE students, 43 control subjects)

Outcome Measured	Desired Effect	Effect of LRE	time-2 mean LRE	time-2 mean Cont?	Agg. Std. Dev.
EFFECTS OUTSIDE THE THEORETICAL MODEL					
1. Factual knowledge of the law and legal processes	+	+12%*	73%	61%	15.75
2. Frequency of telling parents about something useful learned in a class	+	+1.1*	4.1	2.9	3.04
3. Students' rating of this course relative to others taken (better, same, worse)	+	+.5*	2.5	2.0	.63
THEORY-BASED CORRELATES OF BEHAVIOR					
COMMITMENT					
4. Perceived opportunities for demonstrating competence to teachers	+	+.4*	3.9	3.5	.74
5. Importance of doing well and being regarded as a good student in school	+	0	2.7	2.6	.38
6. Importance of doing well and being regarded as a good student in this class	+	+.1	2.6	2.5	.43
7. Students' rating of this course as really helpful	+	+.5*	2.5	2.0	.57
ATTACHMENT					
8. Really liking some teachers and believing they care about you as a person	+	+.4*	3.9	3.5	.82
9. Support offered by teachers to build your interest and help you	+	+.2*	3.4	3.2	.75
INVOLVEMENT					
10. Timely completion of assignments and coming to class prepared to participate	+	+.2*	3.7	3.4	.65
11. Amount of time spent doing homework	+	+.3	2.8	2.6	1.13
12. "Clockwatching" in this class	-	-1.0*	3.0	4.0	1.02
13. Encouragement from the teacher in this class of special projects by students	+	+.6*	3.5	2.9	1.05
BELIEF					
14. Favorable attitudes toward police	+	+.2*	3.5	3.3	.77
15. Belief that judges try to be fair and just	+	0	3.9	3.9	.93
16. Unfavorable attitudes toward deviance	+	0	3.5	3.4	.53
17. Favorable attitudes toward personal violence	-	-.3*	2.2	2.7	.87
18. Rationalizations that delinquent behavior is acceptable sometimes	-	-.2*	2.0	2.2	.70
EQUALITY					
19. Belief that you are treated fairly in school with respect to rules, grades	+	+.5*	4.2	3.7	.76
20. Perception that the rules in this class have applied the same to everybody	+	+.7*	4.2	3.5	.97
21. Perception that the teacher in this class grades fairly	+	+1.0*	4.3	3.3	1.00
LABELING					
22. Your receive praise at home when you do something good at school	+	+.3*	3.7	3.4	1.09
23. Your parents would agree that you get into trouble, are a bad kid	-	-.2*	2.0	2.4	.77
24. Your teachers would agree that you get into trouble, are a bad kid	-	-.1	2.0	2.4	.93
25. Your friends would agree that you get into trouble, are a bad kid	-	+.1	2.1	2.3	1.01
PEER RELATIONSHIPS					
26. Delinquent peer influence	-	0	2.2	2.3	.86
27. Exposure to delinquent peers	-	0	1.7	1.8	.70
28. Students in this class willing to help one another with questions, course work	+	+.2*	3.7	3.5	.83
29. When other students speak in this class, they have something worthwhile to say	+	+.5*	3.8	3.4	.82
30. The other students in this class pay attention when you are talking	+	+.4*	3.8	3.4	.73
FREQUENCIES OF DELINQUENT BEHAVIOR					
31. School rule infractions (cheat on tests, skip school, two more)	-	-2.8*	4.1	7.7	5.49
32. Drinking alcohol	-	-.9*	2.7	3.6	3.50
33. Violence against other students (using a knife, rock, or stick)	-	0	.1	.1	.64
34. Minor theft (steal less than $50, joyride)	-	-.4	1.4	2.1	3.56
35. Vandalism (damage or destroy school or public property)	-	-.7	1.0	2.4	3.06
36. Go out with a group planning to fight or break the law	-	-.5*	.7	1.2	1.86
37. Other status offenses (lie about age, run away)	-	-.3	1.0	1.5	2.20
38. Index offenses (strongarm, break and enter, car theft, two more)	-	0	.6	.8	1.98
39. Minor fraud (avoid paying for food, movies, shows)	-	-.4	.7	1.0	1.76
40. Smoking marijuana	-	0	.8	.7	2.30
41. Hard drug use	-	-.2	.2	.3	1.21

The effects of LRE shown are over and above the possible effects of time-1 score & age.

*Significant at .05 or better (one-tailed test)

toward police (measure #14),* the control students at the end of the fall 1983 semester scored as high on this measure as the LRE students had scored the year before. Since the bulk of ninth-graders were receiving LRE in 1983, word-of-mouth may have resulted in a spread of this favorable effect to students who did not have direct classroom experience with police officers. Also, the LRE classrooms were located close to the control classroom in the same school corridor; all ninth-grade students would have become accustomed to the regular presence of police for reasons that had nothing to do with trouble.

SUMMARY

Neither the suggestive evidence from other sites nor the favorable outcomes at Bill Reed Junior High School imply that adopting an LRE textbook and offering a course by that name will automatically produce desirable effects. The LRE classes studied nationally in 1981 and 1982 varied widely in their levels of success; a few classes even appeared to *worsen* student attitudes and *increase* the frequency of delinquent behavior. Our observation of training and classrooms, logs submitted by teachers, and discussions with the principal indicate that the features associated with effectiveness in reducing delinquency were present to an extracrdinarily high degree at Reed. Our study of LRE classes at other schools suggests strongly that *the major reason LRE was so effective at Bill Reed Junior High School was that these features were a prominent part of the program.* Distinguishing this program from many less successful ones were the extensive use of outside resource persons, the activities used to engage students, the choice and way of presenting illustrative material, and strong administrative support. Based on evidence from the national study, increases in student knowledge probably would have occurred in the absence of other features; improvements in behavior and attitudes probably would not.

---

*Favorable attitudes toward police were signified by agreement with the statements: *Police try to give all kids an even break; Police always have a good reason when they stop somebody;* and *You have a lot of respect for the police in this town.*

ADDENDUM

A PRELIMINARY RESPONSE TO JAMES P. SHAVER'S PAPER, "THE LAW-RELATED EDUCATION EVALUATION PROJECT: A METHODOLOGICAL CRITIQUE OF THE 'IMPACTS ON STUDENTS' FINDINGS"

Grant Johnson
Center for Action Research, Inc.

In light of the arguments presented and the means used to support them, the severe tone of the critique seems out of place. The reader of that paper should be aware that its strong language is not matched by strong arguments. This is true of the review of both years' reports, but becomes especially evident with respect to the more important of the two documents--the Year 2 Law-Related Education Evaluation Report (SSEC-CAR, 1983). This necessarily brief addendum will present some of the more obvious grounds for concluding that the critique does not give cause to revise any conclusion contained in the report.

Selectivity. A defensible critique must comprehend and present accurately the research which it examines. Stating at the outset that "the critique is not meant to be all inclusive" (Shaver, p. 2) does not give license to duck the responsibility of accurate reporting. The conclusions in the review not only rely on selective choices of data at a number of important points, but often convey the impression that some selected subset constitutes the totality of one or another dimension of the study. That the critique highlights weaknesses rather than strengths is understandable. That it repeatedly ignores substantial evidence that would contradict its assertions is inexcusable. Had this been a research paper, rather than a critique of one, many would regard its failure to report contrary evidence as falsification of data.

The problem of misrepresentation in the critique is prominent in its presentation of outcome findings from the evaluation report. Besides examining behavior frequencies (through self-reports, which will be discussed shortly), the evaluators measured twenty-two other relevant outcomes. One of the conclusions stated in the report was that, of these twenty-two, four measures appeared particularly unsusceptible to effects from LRE. As acknowledged in the evaluation report,

> Attitudes toward police were affected favorably more often than unfavorably, but this was not true of attitudes toward deviance, attitudes toward violence, or rationalizations for deviance. The pattern for effects on belief at the Colorado site is similar; there, the data indicate significant favorable impact on attitudes toward police, a nonsignificant favorable effect on attitudes toward personal violence, and no effect on either of the other two belief measures. Among the theoretical antecedents measured, belief appears uncommonly resistant to change from this course of instruction (SSEC-CAR, 1983, pp. 4-49 to 4-50).

In short, the evaluation report itself identified an 18 percent subset of outcome measures (four out of twenty-two) for which findings were notably weak. In its review of findings from the Colorado site, the critique reported only on these four outcomes (Shaver, p. 14) and failed to mention that there were eighteen more, eleven of which showed significant favorable effects. Purposive selection of this weak subset of findings and exclusion of the remaining 82 percent led to a conclusion in the critique that "these results hardly seem to constitute 'strong and defensible findings . . .'" Shaver, p. 14). (Also contributing to this conclusion was an apparent oversight in the critique's reporting of findings pertaining to delinquent behavior at this site. As was called to the reviewer's attention subsequent to printing of the critique, there were no instances--significant or not--of behavioral results "in the wrong direction," although the review contradicts the data in the report by stating that there were four such instances.)

The critique is selective again in reporting outcomes from the national study sites (Shaver, p. 15). The report singled out the ten behavioral categories and fourteen of the other outcome measures as particularly problematic at sites not having strong equivalence between experimental and comparison groups at time-1. The only national site outcomes mentioned in the review are the ones based on these measures. Not mentioned are outcome dimensions that had yielded fifty-nine significant effects favoring the LRE classes and only four favoring the comparison classes. With this evidence excluded from the computation, the critique made it appear that even the modest claim in the report of "suggestive favorable evidence" from the national site was unwarranted.

Worth noting at this point is that in the report all outcome findings for the national site appeared in a single table, as did all outcome findings

for the Colorado site. Selection of a weak subset from either table required extracting individual rows of figures from various parts of the page and ignoring the rows in between those chosen.

A less consequential instance of depiction in the review of a weak subset of findings as the totality occurred through a second apparent oversight (again, the reviewer has been so informed). One section of the report pertains to relationships found between eleven classroom observation variables and effects on delinquent behavior. The report notes that for five of the observation variables, the associations were weak or nonexistent. The highest correlation for any of these five was .19, but correlations between each of the remaining six observation variables and effects on delinquent behavior ranged from .29 to .55 (SSEC-CAR, p. 5-17). Nevertheless, the critique informs us that "the 11 observational variables either were not associated or were only weakly associated with outcomes (with no correlation coefficient greater than .19, apparently)" (Shaver, p. 11).

Self-Reported Delinquency Measures. The critique faults the report repeatedly for omitting the qualifier "self-reports of . . ." when using the terms "delinquency" and "delinquent behavior," as well as for relying on self-reports in the first place (Shaver, pp. 2, 8, 10). Presumably, the qualifier is deemed desirable because delinquency is not a directly observed variable. Although the report makes no secret of the kind of measures used, recurring reference to the method arguably might help persons who read only scattered sentences of the text. The same logic holds as well for every other dimension measured through questionnaire or interview responses, in this and other studies. Attitudes, opinions, exposure to delinquent peers, judgments about the helpfulness of a course, and even knowledge, all are measured through responses, not through direct observation. In the unusual sense that the critique uses it (to indicate technical, rather than grammatical, inaccuracy in the source cited), the parenthetical "(sic)" by this logic would apply equally well to every mention of every outcome variable, not just to delinquency. Instead of prefacing scores of sentences with "according to responses obtained," the report follows the convention of describing the measurement methods used, then relying on the reader not to forget them.

With respect to the appropriateness of relying on self-reports in measuring delinquency, to say that "the validity of self-reports has been . . . and continues to be . . . a concern among sociologists" (Shaver, p. 2) understates the case. The validity of self-reports of delinquency (as well as their reliability and suitability for different types of studies) has been the subject of many major studies during the past twenty years (averaging nearly one a year). Researchers have compared self-reports of delinquent behavior with official police and court records, have employed undercover informants or "tails" to observe subjects' behavior before self-reports are obtained, have administered polygraph tests, and have compared self-reports with reports obtained from victims. This wealth of research has demonstrated repeatedly that self-reports are sufficiently valid and reliable to be suitable by themselves to measure delinquent behavior in field studies.* Of incidental interest, a report of high validity based on confirmation through polygraph tests came from an article (Clark and Tifft, 1966) cited in the critique to substantiate its allegation of "concern among sociologists" (Shaver, p. 2). The task of the LRE evaluation was not to retest the established validity of self-reports; the evaluators accepted the accumulated evidence that this was an acceptable method of measurement.

The critique offers an alternative explanation for findings of reduced or increased self-reported delinquency among students in particular classrooms: "If student-teacher rapport is good; the students may respond in ways to support the teacher's expectations; if rapport is not good, the subjects may respond contrary to what they perceive as the experimenter's desires" (Shaver, p. 3). This contention (actually a theory, which seems out of harmony with the derrogation of theory that appears elsewhere in the critique) suggests that students who like their teacher will under-report their frequencies of behavior, while those who do not like their teacher will report more offenses than they have committed. In the absence of

---

*A summary of studies and conclusions on this topic appears in Hindelang, M., T. Hirschi, and J. Weis, Measuring Delinquency, Beverly Hills: Sage, 1981.

an empirical test of this hypothesis; the following comments bear on its plausibility:

1. Except in the most general terms, students were not informed by the researchers of the design or purpose of the study. Moreover, judging from the kinds of questions they asked during administration of the questionnaires, design and purpose were not topics they cared to know more about.

2. Students received repeated written and verbal assurance of the confidentiality of their responses. The assurance included the information that, if a researcher divulged their answers to a teacher or anyone else, the researcher was subject to a $10,000 fine.

3. With respect to many of the scales in the questionnaire (other than delinquency measures), inventing a pattern of fake responses that would either "meet the expectations of the teacher" or somehow serve to get even with an unliked teacher would have been virtually impossible--particularly with only forty-five or fifty minutes available for responding to more than 150 test items. After answering questions on the rest of the test truthfully, a decision by students to "fake" the delinquency items would have produced glaring inconsistencies. These were not in evidence; for example, students with the highest self-reported delinquency tended to fall at the undesirable end of the range on the other measures.

4. In studies where independent verification of self-reports was possible (in the discussion of validity, above), over-reporting of delinquent behavior has not been found. In a study where the exact number of actual offenses (cheating on tests) was known ahead of time, there was not a single instance of over-reporting.*

5. The probability of over- or under-reporting for any reason would have been the same in the comparision classes as it was in the experimental classes.

Incidentally, the last point suggests a drawback to using official records of delinquency rather than self-reports in an evaluation of law-related education. Some have argued that the course teaches students how to avoid getting caught, even if their level of delinquent behavior stays the same. (In fact, two teachers--not part of the impact study--boasted to this writer that they included lessons on this precise topic in their LRE courses.) To the extent that this argument is true, the bias of arrest or court records would have differed between LRE and comparison students. Had this method of measurement been used, learning to avoid apprehension would have been an alternative explanation for any apparent reduction in delinquent behavior.

---

*Erickson, M., and W. B. Smith, Jr., "On the Relation Between Self-Reported and Actual Deviance: An Empirical Test," Humbolt Journal of Social Relations 2(1974):106-113.

Theory. A widely accepted justification for inferring a causal relationship when a significant correlation is found between two variables is formulation before a study begins of theory-based hypotheses or predictions to be tested. This is what occurred in the LRE evaluation and justifies both the use of one-tailed tests of significance, which the critique finds acceptable (Shaver, p. 15), and the inference of cause from associations found, which the critique rejects (Shaver, pp. 3-4). Treatment in the critique of "theory-testing" pertains not to the theory of delinquency causation upon which design of the study was based, but to a series of exploratory tests of relationships between classroom practice and setting variables and outcomes. The latter were clearly labeled in the report as "after-the-fact," and the report's account of these findings contained a detailed discussion of the caveats required in that kind of analysis. Ignored by the critique, the test of the theory of delinquency causation yielded overwhelmingly supportive findings. As noted earlier, treatment in the critique of the analysis of classroom practices left out the majority of those findings. Once again, selectivity is the basis for an unwarranted conclusion in the critique, this time "inconsistent support of the delinquency theory by the findings of the study" (Shaver, p. 18).

What alarmed this writer was not the treatment in the critique of theory-testing in the LRE study (which paralelled the critique's treatment of several other matters), but comments in that paper about the way educators regard theory in general. Besides referring to the LRE theoretical findings as "distractors" (Shaver, p. 17), the critique notes that "teachers have generally not found the results of theory-oriented research to be useful" and reports "general skepticism about research efforts to build theory and about the usefulness of such theory in educational practice" (Shaver, p. 17). Elsewhere, the critique states that:

> the primary purpose of educational evaluation, as distinct from social science research, is to provide evidence and conclusions in regard to whether an educational practice "works" and under what conditions . . . Questions about theory are secondary . . . (Shaver, p. 16).

Admittedly, my own extended contacts with teachers and school administrators are limited to about a dozen. Either this small sample is atypical among educators, or the implication that educators are not particularly interested in why something works, but only in whether it works is defamatory to those in the profession. I hope the latter is true.

*** 6 ***

Miscellany. Presumably, what is called "educational significance" in the critique refers primarily to information for estimating the proportion of control subjects that an average LRE student would surpass on a given measure. The evaluators routinely obtain this information and intend to include it in the third-year report now in preparation.

With respect to "verification of the independent variable," the observations referred to in the critique as "three times" (Shaver, p. 11) were three two-day observations as stated in the report. The section in the report used to justify a claim in the critique of "inadequate attention to the reliability and validity of the observational data" (Shaver, p. 11) describes six approaches taken to assure validity and several to assure reliability (SSEC-CAR, pp. 3-13 and 3-14). The only shortcoming suggested in that section was lack of resources to send teams of observers instead of individuals into a given class on the same day.

The critique disparages what was termed in the report a "soft" estimate of effects (Shaver, p. 13). As stated in the report, that estimate was undertaken only to enrich the formative evaluation by suggesting further discrimination among classes than were obtained in the regular analysis. This consisted of an examination of raw change scores (not posttest scores only as stated in the critique), and the evaluators labeled the procedure "soft" in recognition of its weaknesses. No findings developed in this manner were used in reporting impact. Incidentally, the critique states that, "in fact, their conservative estimate [of impact, based on the multiple regression procedure] is an estimate of the results that would have been obtained had there been time-1 equivalence (Cohen and Cohen, 1974, Ch. 9)." (Both the quote and the citation are from Shaver, p. 13.) In the chapter cited, Cohen and Cohen assert just the opposite. They call such an interpretation "clearly *not* warranted" (italics theirs, p. 398) and caution that the use of this procedure "does not warrant interpretations cast in the subjunctive form, for example, '*were* the groups equal with regard to the covariates, *then* the differences in the dependent variable would be thus and such" (p. 402).

Fianlly, this writer shares the concern (Shaver, p. 1) that headlines and brief summary accounts of research may mislead. In view of all of the foregoing, however, the capacity of the cited headlines to mislead is trivial compared to the misleading information conveyed by the critique just examined.